SO5032 Lecture 10

Brendan Halpin

April 17, 2023

Outline

SO5032 Lecture 10

SO5032 Lecture 10

Worked example

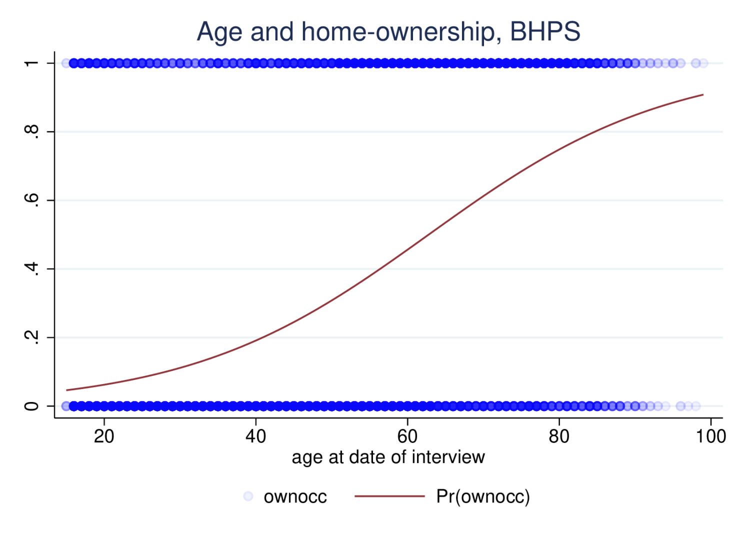
Housing tenure

- Housing tenure: probability of owning outright, BHPS data

```
. logit ownocc age

Iteration 0:   log likelihood = -8728.6773
Iteration 1:   log likelihood = -7150.2389
Iteration 2:   log likelihood = -7095.7194
Iteration 3:   log likelihood = -7095.5268
Iteration 4:   log likelihood = -7095.5268

Logistic regression                                     Number of obs =  14,182
                                                        LR chi2(1)    = 3266.30
                                                        Prob > chi2   =  0.0000
Log likelihood = -7095.5268                             Pseudo R2     =  0.1871

------------------------------------------------------------------------------
      ownocc | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0633183   .0012705    49.84   0.000     .0608281    .0658084
       _cons |  -3.974023   .0697795   -56.95   0.000    -4.110788   -3.837258
------------------------------------------------------------------------------
```

sociology
UNIVERSITY OF LIMERICK

Predictions



sociology
UNIVERSITY OF LIMERICK

Predictions

$LO = a + bX$

$Odds = \exp(a + bX)$

$P = Odds/(1 + Odds)$

X increases by 1:

- LO by b (additive)
- Odds by e^b (multiplicative)
- P is more complicated

Predicton

- Log-odds

$X = x$	$LO(x) = a + bx$
$X = x+1$	$LO(x+1) = a + b(x + 1) = a + bx + b$
Difference:	$LO(x+1) - LO(x) = b$

sociology
UNIVERSITY OF LIMERICK

Prediction: odds scale

- Odds

X = x	Odds(x) = $e^{a+bx} = e^a e^{bx}$
X = x+1	Odds(x+1) = $e^{a+b(x+1)} = e^{a+bx+b} = e^a e^{bx} e^b$
Ratio	Odds(x+1)/Odds(x) = e^b

- Hence odds-ratio: if X increases by 1, OR increases by factor of e^b

sociology
UNIVERSITY OF LIMERICK

Odds ratio

```
. tab univ ownocc
```

univ	ownocc 0	1	Total
0	8,335	3,835	12,170
1	1,514	499	2,013
Total	9,849	4,334	14,183

OR = (499/1514) / (3835/8335) = 0.7163

```
. logit ownocc i.univ

Iteration 0:   log likelihood =  -8729.863
Iteration 1:   log likelihood = -8710.9025
Iteration 2:   log likelihood = -8710.8468
Iteration 3:   log likelihood = -8710.8468

Logistic regression                                     Number of obs = 14,183
                                                        LR chi2(1)    =  38.03
                                                        Prob > chi2   = 0.0000
Log likelihood = -8710.8468                             Pseudo R2     = 0.0022
```

ownocc	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
1.univ	-.3336103	.0551837	-6.05	0.000	-.4417683	-.2254522
_cons	-.7762941	.0195124	-39.78	0.000	-.8145376	-.7380506

$$e^b = e^{-.3336103} = 0.7163$$

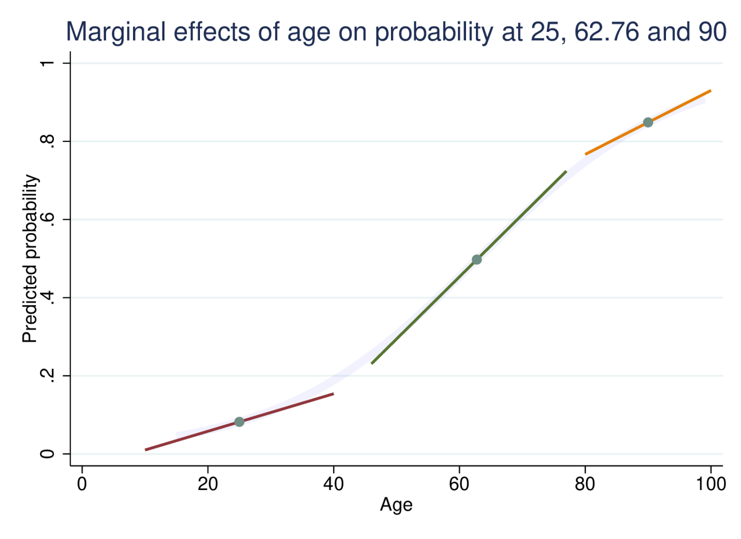
sociology
UNIVERSITY OF LIMERICK

Predictions on probability scale

- Effect of X on the probability scale is non-linear
- Low when p is either high or low
- Highest at $p = 0.5$, odds $= 1$, log-odds $= 0$
- The steepest slope is at $p = 0.5$, with a value of $\frac{\beta}{4}$

sociology
UNIVERSITY OF LIMERICK

Marginal effects



Multiple explanatory variables

```
. logit ownocc age i.univ

Iteration 0:   log likelihood = -8728.6773
Iteration 1:   log likelihood = -7150.3435
Iteration 2:   log likelihood = -7094.4048
Iteration 3:   log likelihood = -7094.1883
Iteration 4:   log likelihood = -7094.1882

Logistic regression                                     Number of obs =  14,182
                                                        LR chi2(2)    = 3268.98
                                                        Prob > chi2   =  0.0000
Log likelihood = -7094.1882                             Pseudo R2     =  0.1873
```

ownocc	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age	.0636471	.0012888	49.38	0.000	.061121	.0661731
1.univ	.0999785	.0608614	1.64	0.100	-.0193076	.2192646
_cons	-4.004807	.0724889	-55.25	0.000	-4.146883	-3.862731

sociology UNIVERSITY OF LIMERICK

SO5032 Lecture 10

Inference

Inference

- In practice, inference is similar to OLS though based on a different logic
- For each explanatory variable, $H_0 : \beta = 0$ is the interesting null
- $z = \frac{\hat{\beta}}{SE}$ is approximately normally distributed (large sample property)
- More usually, the Wald test is used: $\left(\frac{\hat{\beta}}{SE}\right)^2$ has a χ^2 distribution with one degree of freedom

sociology
UNIVERSITY OF LIMERICK

Likelihood ratio tests

- The "likelihood ratio" test is thought more robust than the Wald test for smaller samples
- Where l_0 is the likelihood of the model without X_j, and l_1 that with it, the quantity

$$-2\left(\log\frac{l_0}{l_1}\right) = -2\left(\log l_0 - \log l_1\right)$$

is χ^2 distributed with one degree of freedom

sociology
UNIVERSITY OF LIMERICK

Nested models

- More generally, $-2\left(\log \frac{l_0}{l_1}\right)$ tests nested models: where model 1 contains all the variables in model 0, plus m extra ones, it tests the null that all the extra β coefficients are zero (χ^2 with m df)
- If we compare a model against the null model (no explanatory variables, it tests

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

- Strong analogy with F test in OLS

Example

```
. qui logit ownocc age

. est store mod1

. logit ownocc age i.educ

Iteration 0:   log likelihood = -8728.6773
Iteration 1:   log likelihood = -7136.2054
Iteration 2:   log likelihood = -7077.7722
Iteration 3:   log likelihood = -7077.5203
Iteration 4:   log likelihood = -7077.5203

Logistic regression                                     Number of obs =  14,182
                                                        LR chi2(3)    = 3302.31
                                                        Prob > chi2   =  0.0000
Log likelihood = -7077.5203                             Pseudo R2     =  0.1892

------------------------------------------------------------------------------
      ownocc | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0652599   .0013433    48.58   0.000     .0626271    .0678927
             |
        educ |
        Med  |   .3041599   .0673504     4.52   0.000     .1721556    .4361642
         Lo  |  -.1075582   .0461399    -2.33   0.020    -.1979907   -.0171257
             |
       _cons |  -4.060514   .0730524   -55.58   0.000    -4.203694   -3.917333
------------------------------------------------------------------------------

. lrtest mod1

Likelihood-ratio test
Assumption: mod1 nested within .

 LR chi2(2) =  36.01
Prob > chi2 = 0.0000
```

SO5032 Lecture 10

Margins command

"Average Marginal Effect"

- "What would happen to the averege predicted probability if we increased X?"
- For linear regression, increase X by 1 => increase by b
  - increase X by 10 => increase by b× 10
  - increase X by 0.1 => increase by b× 0.1
  - since it's a straight line
- For AME in logistic we use the slope of the tangent, for each X value
- Average across the observed data
- Gives something like a LPM slope

sociology
UNIVERSITY OF LIMERICK

AME in Stata

```
. margins, dydx(age)

Average marginal effects                         Number of obs = 14,182
Model VCE: OIM

Expression: Pr(ownocc), predict()
dy/dx wrt:  age

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0104836   .0001382    75.84   0.000     .0102126    .0107545
------------------------------------------------------------------------------
```

sociology
UNIVERSITY OF LIMERICK

SO5032 Lecture 10

Maximum likelihood

Maximum likelihood estimation

- What is this “likelihood”?
- Unlike OLS, logistic regression (and many, many other models) are extimated by *maximum likelihood estimation*
- In general this works by choosing values for the parameter estimates which maximise the probability (likelihood) of observing the actual data
- OLS can be ML estimated, and yields exactly the same results

Iterative search

- Sometimes the values can be chosen analytically
  - A likelihood function is written, defining the probability of observing the actual data given parameter estimates
  - Differential calculus derives the values of the parameters that maximise the likelihood, for a given data set
- Often, such "closed form solutions" are not possible, and the values for the parameters are chosen by a systematic computerised search (multiple iterations)
- Extremely flexible, allows estimation of a vast range of complex models within a single framework

Likelihood as a quantity

- Either way, a given model yields a specific maximum likelihood for a give data set
- This is a probability, henced bounded $[0:1]$
- Reported as log-likelihood, hence bounded $[-\infty:0]$
- Thus is usually a large negative number
- Where an iterative solution is used, likelihood at each stage is usually reported – *normally* getting nearer 0 at each step

SO5032 Lecture 10

Tabular data

Tabular data

- If all the explanatory variables are categorical (or have few fixed values) your data set can be represented as a table
- If we think of it as a table where each cell contains n yeses and $m - n$ noes (n successes out of m trials) we can fit grouped logistic regression
- n successes out of m trials implies a binomial distribution of degree m

$$\log \frac{n}{m-n} = \alpha + \beta X$$

- The parameter estimates will be exactly the same as if the data were treated individually

Tabular data and goodness of fit

- But unlike with individual data, we can calculate goodness of fit, by relating observed successes to predicted in each cell
- If these are close we cannot reject the null hypothesis that the model is incorrect (i.e., you want a high p-value)
- Where l_i is the likelihood of the current model, and l_s is the likelihood of the "saturated model" the test statistic is

$$-2\left(\log \frac{l_i}{l_s}\right)$$

- The saturated model predicts perfectly and has as many parameters as there are "settings" (cells in the table)
- The test has *df* of number of settings less number of parameters estimated, and is χ^2 distributed